

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/139923/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Irvine, Elizabeth ORCID: <https://orcid.org/0000-0003-1861-9533> 2021.  
Assessing unlimited associative learning as a transition marker. *Biology and Philosophy* 36 , 21. 10.1007/s10539-021-09796-0 file

Publishers page: <https://doi.org/10.1007/s10539-021-09796-0>  
<<https://doi.org/10.1007/s10539-021-09796-0>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





# Assessing unlimited associative learning as a transition marker

## Commentary on Birch et al. 2020, Unlimited Associative Learning and the Origins of Consciousness: A Primer and Some Predictions

Elizabeth Irvine<sup>1</sup> 

Received: 25 January 2021 / Accepted: 8 March 2021  
© The Author(s) 2021

### Abstract

The target paper (building on Ginsburg and Jablonka in JTB 381:55–60, 2015, The evolution of the sensitive soul: Learning and the origins of consciousness, MIT Press, USA, 2019) makes a significant and novel claim: that positive cases of non-human consciousness can be identified via the capacity of unlimited associative learning (UAL). In turn, this claim is generated by a novel methodology, which is that of identifying an evolutionary ‘transition marker’, which is claimed to have theoretical and empirical advantages over other approaches. In this commentary I argue that UAL does not function as a successful transition marker (as defined by the authors), and has internal problems of its own. However, I conclude that it is still a very productive anchor for new research on the evolution of consciousness.

**Keywords** Transition marker · Evolution of consciousness · Associative learning

*A different methodology:* A fairly standard way of generating empirical markers of consciousness is to posit some capacity that is identical to or highly correlated with the presence of consciousness, and then generate operationalisations of that capacity. So, if one thinks that global broadcast of information in the brain is what makes content conscious, then one searches for ways to operationally test for that, for example by tracking neurophysiological activity, or specific behaviours that are only possible as a result of global broadcast.

According to the target paper, investigating consciousness using transition markers works differently. Transition markers are supposed to be based on an agreed set of sufficient criteria for the state in question. Even if there are disagreements about which criteria are necessary for something to be classed as (for example)

---

✉ Elizabeth Irvine  
irvinee@cardiff.ac.uk

<sup>1</sup> School of English, Communication and Philosophy, Cardiff University, Cardiff, Wales

alive or conscious, or if there are differences in theories about life or consciousness, researchers can usually agree on what counts as a set of sufficient conditions. Taking these, one can try to identify a single marker that requires that all of these sufficient conditions be satisfied. Any entity that displays this single marker therefore satisfies all the sufficient conditions for being in a particular state.

Other aspects of the idea of transition markers trace back to Gánti (2003) and Maynard Smith and Szathmáry's (1997) work on the evolution of life from non-living systems. Gánti proposed an abstract, dynamical chemical model of a 'chemoton' which satisfied his (sufficient) criteria for living systems. In at least some forms, the chemoton was capable of 'unlimited heredity': it was capable of open-ended evolution, which is a feature of all living systems. Changing some of the features of the chemoton makes it capable of only limited heredity, associated with being a 'minimal' living system. These different versions of the model therefore suggest a possible evolutionary path from non-living systems, to minimal living systems with limited heredity, to 'full' living systems with unlimited heredity. Unlimited heredity functions as a 'transition marker' in this evolutionary path to living systems.

Birch et al. (also Ginsburg and Jablonka 2015, 2019) argue that focussing on a single transition marker, rather than a longer list of criteria, comes with a range of methodological advantages. They state that using transition markers "...allows theoretical and empirical research programmes to aim at a single, common goal, despite a potentially significant amount of underlying divergence in views about the fundamental nature of [the concept in question]" [p. 5 of 24 in ms.]. Further, the abstract nature of transition markers allows researchers to construct minimal models of e.g. living or conscious systems, to identify transitional evolutionary states between e.g. non-conscious and conscious organisms, to date the emergence of these states, and (plausibly) to map the transition marker to different realisations across different neurobiological and neurophysiological systems.

On these grounds Ginsburg and Jablonka (2015) criticise the standard approach to investigating (non-human) consciousness as essentially failing to offer the above advantages, and for failing to "go beyond a re-description of some of the necessary conditions for [the occurrence of consciousness]" (*op cit*, p. 58). So, for example, if one thinks that consciousness requires the global broadcast of information, global workspace models of consciousness merely re-describe and offer straightforward operationalisations of this necessary condition. Where researchers disagree on necessary conditions (as they do), these operationalisations are not helpful for unifying and driving research, and neither does this approach suggest how the capacity evolved.

I think these criticisms are reasonable ones, but I next argue that the transition marker identified for consciousness in the target paper, unlimited associative learning (UAL), does not necessarily escape these problems.

*UAL as a partial redescription:* The criteria deemed sufficient for consciousness by Birch et al. are roughly these: global accessibility, feature binding, attention, intentionality, integration of information over time, evaluative system(s) and agency, and recognition of (embodied) self vs. other. For unlimited associative learning (UAL) to be exhibited, it is claimed that these other markers must be satisfied: UAL therefore functions as a transition marker of consciousness. UAL is differentiated

from *limited* associative learning in specific ways. UAL requires the capacities for associative learning of compound and novel stimuli, second-order and trace conditioning, and flexible and easily overwritten reward learning.

The first issue is that aside from calling these capacities sufficient rather than necessary conditions for consciousness, UAL is partly just a redescription and straightforward operationalisation of some of these conditions. For example, associative learning of compound stimuli (part of UAL) is a fairly plausible operationalisation of the sufficient condition of feature binding, and UAL of multi-sensory stimuli is a plausible operationalisation of some form of global broadcast. Trace conditioning (again, part of UAL) is a way to operationalise the sufficient condition of integration of information over time. Reward learning is a straightforward operationalisation of the sufficient condition of having an evaluative system. Attention (another sufficient condition) is plausibly linked to UAL, but is perhaps subsumed by the criterion of global broadcasting (typically, global broadcast is a feature of only attentionally selected inputs). Finally, associative learning of novel stimuli seems to partly redescribe the ability for (unlimited) associative learning in the first place. In this case, UAL is subject to the same criticisms of e.g. the global workspace model in merely redescribing and operationalising an existing set of criteria (see Table 1 for summary).

*Another transition marker:* The second issue is that UAL does not obviously connect to some of the other conditions identified as sufficient for consciousness, or at least is not the best way to capture these conditions in a transition marker. In Birch et al., intentionality is said to be required for representing and storing learned associations. The target article also suggests that agency and embodiment are required for (any form of) action selection, and that these combined with self-other registration are necessary for (any form of) learning about the world (pp. 12–13 of ms.). These claims are hard to evaluate given their brief description in the paper, but their Table 1 describes suggested associated mechanisms and behavioural signatures. From here, these conditions (intentionality, agency, embodiment, self-other registration) essentially turn out to refer to the capacity of an organism to generate egocentric representations of itself acting in space, where actions are goal-directed and selected in a top-down manner. This package of capacities has been separately defended as sufficient for consciousness by Merker (2007) and Barron and Klein (2016).

Arguably, the package of abilities linked with egocentric representations of the moving body in space provides a much better transition marker for tracking the

**Table 1** Summary of relationships between conditions and operationalisations

Sufficient condition for consciousness	Operationalised as a type of associative learning (AL)
Feature binding	AL of compound stimuli
Global broadcast (usually includes attention)	AL of multi-sensory stimuli
Integration of information over time	Trace conditioning
Having an evaluative system	Reward learning

emergence of this second subset of sufficient conditions for consciousness than anything based on unlimited associative learning. UAL is not well placed to capture what is really required for top-down control of goal-directed actions, or representations of the body in the world. They are clearly related: some kinds of associated learning (such as spatial learning) are plausibly only possible when an organism has an egocentric representation of itself in space. However, top-down control of actions is (plausibly) not well tracked by capacities for associative learning. In this case, it makes sense to use a different transition marker, or at least a different package of capacities, to track these criteria.

At this point then, Birch et al.'s sufficient criteria for consciousness split into two: one set that come with fairly straightforward operationalisations in terms of associative learning, and so tracking standard methods for investigating consciousness (see Table 1 again), and another set based on agency, embodiment, and so on that is better captured by a different transition marker or package of related capacities. This is obviously rather messier than the original proposal.

*Does it matter?:* However, I don't think that this splintering up of transition markers is necessarily a problem for research programmes focussed on the evolution of consciousness. UAL gets something right: it suggests that by focussing on different types of learning we can also test for the presence of lots of other cognitive abilities (i.e. those required for those types of learning), that are generally agreed to be sufficient for consciousness. As experimental paradigms for different types of learning are well established, and in many cases have been applied to a range of organisms, they provide a productive focus for research whether UAL is really based on a new method or not.

Similarly, it is not the end of the world if some sufficient conditions for consciousness are better tracked by other means, for example by assessing decision making capacities. Associative learning can certainly inform decision making, but associative learning abilities alone are unlikely to be the best route to tracking the (reasonably) complex decision making capacities associated with the second subset of sufficient conditions. For example, testing associative learning capacities is a perfectly reasonable way of testing whether an organism can integrate information over time, via trace conditioning, but does not seem a particularly direct or effective way of testing whether an organism can engage in online action selection in a complex body in the face of changing sensory information. Instead, it seems plausible to assess these abilities using established behavioural paradigms specific to these abilities, and related neuroanatomical research concerning online action control. There is no obvious theoretical pay-off in cutting connections to relevant areas of empirical research.

In sum then, the target paper tries to provide a focus for empirical and theoretical research by identifying a single transition marker. I have suggested that UAL does not work as a transition marker in the way the authors suggest. However, even if a single transition marker of consciousness cannot be found, there is still unity to be found in the set of agreed sufficient conditions for the presence of consciousness. If numerous research programs emerge from this, each of which target a different set of sufficient conditions, and different sets of mechanisms, it is not clear to me that anything important will have been lost. Indeed, this might open up our ideas further

on there being different evolutionary routes to consciousness, and perhaps different types of consciousness altogether.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Barron AB, Klein C (2016) What insects can tell us about the origins of consciousness. *Proc Natl Acad Sci* 113(18):4900–4908
- Gánti T (2003) *The principles of life*. Oxford University Press
- Ginsburg S, Jablonka E (2015) The teleological transitions in evolution: a Gántian view. *J Theor Biol* 381:55–60
- Ginsburg S, Jablonka E (2019) *The evolution of the sensitive soul: learning and the origins of consciousness*. MIT Press, USA
- Merker B (2007) Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav Brain Sci* 30(1):63–81
- Smith JM, Szathmari E (1997) *The major transitions in evolution*. Oxford University Press

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.